From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy

Year: 2023 | Citations: 523 | Authors: Maanak Gupta, Charankumar Akiri, Kshitiz Aryal, Elisabeth Parker, Lopamudra Praharaj

---

Abstract

Undoubtedly, the evolution of Generative AI (GenAI) models has been the highlight of digital transformation in the year 2022. As the different GenAI models like ChatGPT and Google Bard continue to foster their complexity and capability, it's critical to understand its consequences from a cybersecurity perspective. Several instances recently have demonstrated the use of GenAI tools in both the defensive and offensive side of cybersecurity, and focusing on the social, ethical and privacy implications this technology possesses. This research paper highlights the limitations, challenges, potential risks, and opportunities of GenAI in the domain of cybersecurity and privacy. The work presents the vulnerabilities of ChatGPT, which can be exploited by malicious users to exfiltrate malicious information bypassing the ethical constraints on the model. This paper demonstrates successful example attacks like Jailbreaks, reverse psychology, and prompt injection attacks on the ChatGPT. The paper also investigates how cyber offenders can use the GenAI tools in developing cyber attacks, and explore the scenarios where ChatGPT can be used by adversaries to create social engineering attacks, phishing attacks, automated hacking, attack payload generation, malware creation, and polymorphic malware. This paper then examines defense techniques and uses GenAI tools to improve security measures, including cyber defense automation, reporting, threat intelligence, secure code generation and detection, attack identification, developing ethical guidelines, incidence response plans, and malware detection. We will also discuss the social, legal, and ethical implications of ChatGPT. In conclusion, the paper highlights open challenges and future directions to make this GenAI secure, safe, trustworthy, and ethical as the community understands its cybersecurity impacts.